AN ITERATIVE SPATIO-TEMPORAL SPEECH ENHANCEMENT ALGORITHM FOR MICROPHONE ARRAYS

Malay Gupta and Scott C. Douglas

Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275 USA

ABSTRACT

We present a new spatio-temporal algorithm for speech enhancement using microphone arrays. Our technique involves the development of an iterative method for computing the generalized eigenvectors of the multichannel data as measured from the microphone array. Coefficient adaptation is performed using the spatio-temporal correlation coefficient sequences of the observed data. The technique avoids large matrix-vector multiplications; hence, the computation time and computational resource requirements are significantly smaller than those of competing methods. The technique is applicable to a wide variety of noise types, including stationary correlated noise and non-stationary speech-like (babble) background noise, without requiring any noise-dependent parameter settings.

Index Terms— Speech enhancement, acoustic arrays, adaptive arrays, eigenvalues and eigenfunctions, decorrelation

1. INTRODUCTION

Signal-subspace-based algorithms [1] serve as an alternative to spectral-subtraction-based algorithms [2] for speech enhancement. One challenge with spectral subtraction as used for speech enhancement is the distortion imposed on the enhanced speech in the form of musical tone artifacts. Subspace techniques are based either on the singular value decomposition (SVD) or the eigenvalue decomposition (EVD) of second-order statistics of the observed noisy speech and/or any noise processes corrupting the speech. A key idea in subspace techniques is the decomposition of the noisy speech signal space into two mutually orthogonal subspaces: the signal-plus-noise subspace, and the noise-only subspace. Speech enhancement is performed by nulling signals within the noise subspace and enhancing signals within the signal-plus-noise subspace. This enhancement is possible because speech often fits a low-rank linear model; moreover, subspace-based methods tend to work best for uncorrelated additive white noise interference. While the low-rank linear model for speech is often accurate, the noise corrupting the speech is rarely uncorrelated in practical scenarios. As a result, subspace algorithms typically have lower performance in the presence of correlated noise interference.

Several subspace algorithms designed for single-channel speech enhancement with correlated noise interference have recently been proposed [3, 4, 5, 6]. These methods use a combination of voice activity detection, generalized SVD (GSVD) or generalized EVD (GEVD) processing, and/or spectral domain manipulations to enhance speech corrupted by noise. The techniques in [3, 4] assume that the eigenvectors of the noise are identical to the clean speech eigenvectors, an assumption which is true only for the white noise case. The techniques in [5, 6] integrate noise prewhitening within the algorithms and hence provide better speech enhancement in correlated noise scenarios.

Microphone-array-based signal processing techniques [7] have recently attracted much interest in the speech enhancement community due to their ability to combine beamforming [8] with the temporal processing of speech for more effective signal enhancement. A multi-microphone subspace algorithm based on the GSVD has been recently proposed in [9] and is an extension of the method in [5]. Computing the GSVD is non-trivial and requires specialized algorithms [10, 11]. Similarly, computing the GEVD also requires more work as compared to the EVD due to an integrated whitening procedure for the former method. Hence, extensions of the methods like [6] to the multi-microphone case are computationally expensive, particularly given the dimensional increases in the calculations due to multi-sensor data sets. For example, in an n-microphone system with an L-tap long filter per channel, the direct computation of the GEVD on an $nL \times nL$ correlation matrix can be difficult for even a few microphones and typical window sizes, such that most existing methods are practical only for short data processing windows (*e.g.* 20 to 80 samples long) that limit their overall effectiveness.

In this paper, we develop a new multi-microphone speech enhancement technique based on an iterative methodology to compute generalized eigenfilters for enhancing speech from spatio-temporal correlation coefficient sequences. The method requires measurements of the noise-only signal field as heard at the microphones and uses a clever time-domain filter update for performing joint diagonalization of the spatio-temporal correlation statistics of both the noisy speech and the background noise signals. The advantage of this technique is in its use of a single eigenfilter for representing an entire nL-dimensional signal subspace by time shifts of the corresponding filter impulse response. Our technique does not involve large matrix-vector multiplications or any matrix inversions and hence is computationally attractive for real-time processing. It also does not require a calibrated microphone array. Application of the method to microphone array data in a laboratory environment indicate that the procedure can achieve significant gains in signal-to-interference ratios (SIRs) even in low SIR environments, without introducing musical tone artifacts in the enhanced speech.

2. SPATIO-TEMPORAL EIGENFILTERING

Let $s(l)$ denote a clean speech source signal which is measured at the output of an n-microphone array in the presence of correlated noise $v(l)$ at time instant l. The output of the j^{th} microphone at time instant l can be written as

$$y_j(l) = v_j(l) + \sum_{p=-\infty}^{\infty} h_{jp} s(l-p) = v_j(l) + x_j(l), \qquad (1)$$

where $\{h_{jp}\}$ are the coefficients of the time-invariant acoustic impulse response between the speech source and the j^{th} microphone. The $x_j(l)$ and $v_j(l)$ signals represent the filtered speech and the noise component at the output of the j^{th} microphone, respectively. The additive noise $v_j(l)$ is assumed to be uncorrelated with the clean speech and has an unknown correlation structure. A vector model incorporating signals from all microphones can be written as

$$\mathbf{y}(l) = \mathbf{x}(l) + \mathbf{v}(l), \tag{2}$$

where $\mathbf{y}(l) = [y_1(l) \ \cdots \ y_n(l)]^T$, $\mathbf{x}(l)$, and $\mathbf{v}(l)$ are n-dimensional vectors corresponding to the observed signal, the clean speech signal and the noise signal at each microphone, respectively.

In this paper, we transform the speech enhancement problem into an equivalent iterative multichannel filtering task in which the multichannel filter output at iteration k is given by

$$\mathbf{z}_k(l) = \sum_{p=0}^{L} \mathbf{W}_p(k)\mathbf{y}(l-p), \tag{3}$$

where the $(n \times n)$ matrix sequence $\{\mathbf{W}_p(k)\}$, $0 \le p \le L$, contains the coefficients of the multichannel adaptive filter at iteration k. For ease of notation, we constrain L to be even-valued. For speech enhancement, we adapt the $\{\mathbf{W}_p(k)\}$ sequence such that the total SIR of the multichannel signal outputs $\mathbf{z}_k(l)$ are maximized. The problem of SIR maximization under correlated noise interference is closely related to calculation of the GEVD and is sometimes referred to as *oriented principal component analysis* (OPCA) [12]. OPCA solves for generalized eigenvectors that, when applied to the data, maximize signal variance and minimize the noise variance for any noise type. We express the total power in the elements of $\mathbf{z}_k(l)$ as

$$\begin{aligned} P(k) &= \text{tr}\left\{ \frac{1}{N} \sum_{l=N(k-1)+1}^{Nk} \mathbf{z}_k(l)\mathbf{z}_k^T(l) \right\} \\ &= \sum_{p=0}^{L}\sum_{q=0}^{L} \text{tr}\left\{ \mathbf{W}_p(k)\mathbf{R}_{\mathrm{y}_{q-p}}\mathbf{W}_q^T(k) \right\}, \end{aligned} \tag{4}$$

where N is the length of the data sequence and tr$\{.\}$ corresponds to the matrix trace. The sequence $\{\mathbf{R}_{\mathrm{y}_p}\}$ denotes the multichannel autocorrelation sequence of $\mathbf{y}(l)$ and is defined as

$$\mathbf{R}_{\mathrm{y}_p} = \frac{1}{N} \sum_{l=N(k-1)+1}^{Nk} \mathbf{y}(l)\mathbf{y}^T(l-p), \quad -\frac{L}{2} \le p \le \frac{L}{2}. \tag{5}$$

In the above equations, the filter $\{\mathbf{W}_p(k)\}$ is zero outside of the range $0 \le p \le L$, and $\{\mathbf{R}_{\mathrm{y}_p}\}$ is constrained to be zero outside of the range $|p| \le (L/2)$. By substituting (2) in (3), we obtain

$$\mathbf{z}_k(l) = \sum_{p=0}^{L} \mathbf{W}_p(k)\mathbf{v}(l-p) + \sum_{p=0}^{L} \mathbf{W}_p(k)\mathbf{x}(l-p). \tag{6}$$

Assuming that the speech and noise signals are uncorrelated with each other, the total output signal power can be written as $P(k) = P_x(k) + P_v(k)$, where

$$P_x(k) = \sum_{p=0}^{L}\sum_{q=0}^{L} \text{tr}\left\{ \mathbf{W}_p(k)\mathbf{R}_{\mathrm{X}_{q-p}}\mathbf{W}_q^T(k) \right\} \tag{7}$$

$$P_v(k) = \sum_{p=0}^{L}\sum_{q=0}^{L} \text{tr}\left\{ \mathbf{W}_p(k)\mathbf{R}_{\mathrm{V}_{q-p}}\mathbf{W}_q^T(k) \right\}, \tag{8}$$

and $\{\mathbf{R}_{\mathrm{X}_p}\}$ and $\{\mathbf{R}_{\mathrm{V}_p}\}$ are the multichannel autocorrelation coefficient sequences of the speech and noise signals, respectively, at the sensors.

In practice, $\{\mathbf{R}_{\mathrm{X}_p}\}$ and $\{\mathbf{R}_{\mathrm{V}_p}\}$ are not directly available; however, we assume that there exists speech silence periods whereby $\{\mathbf{R}_{\mathrm{V}_p}\}$ can be estimated from the sensors using data during these silence periods. Let the number of available noise sequence snapshots be $N_v (<< N)$, such that

$$\mathbf{R}_{\mathrm{V}_p} = \frac{1}{N_v} \sum_{l=N_v(k-1)+1}^{N_v k} \mathbf{v}(l)\mathbf{v}^T(l-p), \quad -\frac{L}{2} \le p \le \frac{L}{2}. \tag{9}$$

As we also do not have access to $\{\mathbf{R}_{\mathrm{X}_p}\}$ to compute $P_x(k)$, we replace $P_x(k)$ by $P(k)$ which is an estimate of the total speech signal power that depends on $\{\mathbf{R}_{\mathrm{y}_p}\}$. Thus, we propose to find the coefficient sequence $\{\mathbf{W}_p(k)\}$ that maximizes the following metric:

$$J(\{\mathbf{W}_p(k)\}) = \frac{\text{tr}\left\{ \sum_{p=0}^{L}\sum_{q=0}^{L} \mathbf{W}_p(k)\mathbf{R}_{\mathrm{y}_{q-p}}\mathbf{W}_q^T(k) \right\}}{\text{tr}\left\{ \sum_{p=0}^{L}\sum_{q=0}^{L} \mathbf{W}_p(k)\mathbf{R}_{\mathrm{V}_{q-p}}\mathbf{W}_q^T(k) \right\}}. \tag{10}$$

The function $J(\{\mathbf{W}_p(k)\})$ is a spatio-temporal extension of the *Rayleigh quotient* [11], such that the sequence $\{\mathbf{W}_p(k)\}$ that maximizes (10) corresponds to the generalized eigenfilters of the multichannel autocorrelation sequence pair $(\{\mathbf{R}_{\mathrm{y}_p}\}, \{\mathbf{R}_{\mathrm{V}_p}\})$. Hence, at the stationary point of (10), the sequence $\{\mathbf{W}_p(k)\}$ satisfies

$$\sum_{p=0}^{L}\sum_{q=0}^{L} \mathbf{W}_p(k)\mathbf{R}_{\mathrm{y}_{q-p}}\mathbf{W}_q^T(k) = \begin{cases} \mathbf{\Lambda} & \text{if } |q-p| = 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$\sum_{p=0}^{L}\sum_{q=0}^{L} \mathbf{W}_p(k)\mathbf{R}_{\mathrm{V}_{q-p}}\mathbf{W}_q^T(k) = \begin{cases} \mathbf{I} & \text{if } |q-p| = 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $\mathbf{\Lambda}$ and $\{\mathbf{W}_p(k)\}$ denote the generalized eigenvalues and eigenfilters of $(\{\mathbf{R}_{\mathrm{y}_p}\}, \{\mathbf{R}_{\mathrm{V}_p}\})$. In other words, the coefficient sequence $\{\mathbf{W}_p(k)\}$ simultaneously diagonalizes the sequences $\{\mathbf{R}_{\mathrm{y}_p}\}$ and $\{\mathbf{R}_{\mathrm{V}_p}\}$. We now describe an algorithm that attempts to solve (11)–(12) in an iterative fashion.

3. AN ALGORITHM FOR SPATIO-TEMPORAL GENERALIZED EIGENVALUE DECOMPOSITION

The algorithm we propose for solving (11)–(12) is inspired by the family of adaptive algorithms presented in [13] and in fact can be viewed as a non-trivial extension of the adaptive EVD method of [13] to the space of multichannel filters. Our algorithm uses spatio-temporal correlation coefficient sequences of dimension $(n \times nL)$ in contrast to other techniques which require $(nL \times nL)$ autocorrelation matrices, thus making our techniques more computationally-attractive.

Define the following multichannel convolution operations involving the coefficient sequence $\{\mathbf{W}_p(k)\}$:

$$\overline{\mathbf{R}\mathrm{y}}_q(k) = \begin{cases} \sum_{p=0}^{L} \mathcal{H}(\mathbf{R}_{\mathrm{y}_{q-p}})\mathbf{W}_p^T(k) & \text{if } -\frac{L}{2} \le q \le \frac{L}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

$$\mathbf{G}\mathrm{y}_p(k) = \begin{cases} \sum_{q=0}^{L} \mathbf{W}_q(k)\overline{\mathbf{R}\mathrm{y}}_{p-q}(k) & \text{if } 0 \le p \le L \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

$$\overline{\mathbf{R}\mathrm{v}}_q(k) = \begin{cases} \sum_{p=0}^{L} \mathcal{H}(\mathbf{R}\mathrm{v}_{q-p})\mathbf{W}_p^T(k) & \text{if } -\frac{L}{2} \le q \le \frac{L}{2} \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

$$\mathbf{G}\mathrm{v}_p(k) = \begin{cases} \sum_{q=0}^{L} \mathbf{W}_q(k)\overline{\mathbf{R}\mathrm{v}}_{p-q}(k) & \text{if } 0 \le p \le L \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

In the above set of equations, $\mathcal{H}(\cdot)$ is a windowing operation given by $\mathcal{H}(\mathbf{R}_p) = h_p\mathbf{R}_p$, where h_p, $-L/2 \le p \le L/2$ is a centered Bartlett window. This windowing operation was found to be necessary to ensure the validity of the estimated autocorrelation sequences and to allow the algorithm to converge to a stable stationary point.

We define an update term of the form

$$\mathbf{G}_p(k) = \frac{f_2(k)}{f_1(k)}\,\overline{\mathrm{triu}}[\mathbf{G}\mathrm{y}_p(k)] + \mathrm{tril}[\mathbf{G}\mathrm{v}_p(k)], \quad (17)$$

where $\overline{\mathrm{triu}}[\mathbf{G}]$ denotes the strictly upper triangular part and $\mathrm{tril}[\mathbf{G}]$ denotes the lower triangular part of the matrix $\mathbf{G}$. The terms $f_2(k)$ and $f_1(k)$ are scaling factors used to adjust the average magnitude of $\{\mathbf{G}\mathrm{y}_p(k)\}$ to match that of $\{\mathbf{G}\mathrm{v}_p(k)\}$ and are defined as

$$f_2(k) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p=0}^{L}|g_{ijp}^{v}(k)| \quad (18)$$

$$f_1(k) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p=0}^{L}|g_{ijp}^{y}(k)|, \quad (19)$$

where $g_{ijp}^{y}(k)$ and $g_{ijp}^{v}(k)$ are the elements of the matrix sequences $\{\mathbf{G}\mathrm{y}_p(k)\}$ and $\{\mathbf{G}\mathrm{v}_p(k)\}$, respectively. Note that if (11)–(12) are satisfied, $\mathbf{G}_p(k) = \delta_p\mathbf{I}$. Finally, we define a correction term for the coefficient updates as

$$\mathbf{U}_p(k) = \sum_{q=0}^{L}\mathcal{H}(\mathbf{G}_{p-q}(k))\mathbf{W}_q(k),\ 0 \le p \le L \quad (20)$$

The coefficient updates are then given by

$$\mathbf{W}_p(k+1) = (1+\mu)c(k)\mathbf{W}_p(k) - \mu\frac{c(k)}{d(k)}\mathbf{U}_p(k), \quad (21)$$

where $d(k) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p=0}^{L}|g_{ijp}(k)|$, and $c(k) = \frac{1}{\sqrt{d(k)}}$ are the scaling factors chosen to stabilize the algorithm and reduce the sensitivity of the update to the chosen step size μ, similar to the algorithm in [14]. Typically, step sizes in the range $0.35 \le \mu \le 0.5$ are chosen and appear to work well.

It can be shown that a stationary point of (21) corresponds to the solution in (11)–(12), such that $\mathbf{G}_p = \delta_p\mathbf{I}$. Extensive simulations indicate that this algorithm achieves this stationary point for typical data sets, and the convergence speed of the algorithm is similar to that of [14]. To illustrate this behavior, Fig. 1 shows plots of $\mathbf{G}\mathrm{y}_p$, $\mathbf{G}\mathrm{v}_p$, and $\mathbf{G}_p$ after 100 iterations of the algorithm operating on single-talker data taken from a three-microphone laboratory setup under babble noise conditions with an initial SIR of -10dB and $L = 512$. Each of the nine plots shows a $g_{ijp}(100)$ for $1 \le j \le 3$ and $0 \le p \le 512$ such that the x-axis of each plot is $n = 512j + p$. The last column of plots shows the entire $\mathbf{G}_p(100)$ sequence, indicating that $\mathbf{G}_p(100) \approx \delta_p\mathbf{I}$.

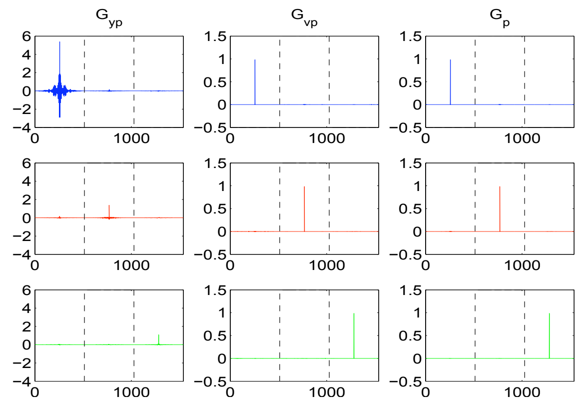

Fig. 1. The sequences $\mathbf{G}\mathrm{y}_p$, $\mathbf{G}\mathrm{v}_p$, and $\mathbf{G}_p$ obtained after 100 iterations of (21) as applied to three-microphone data; see text for explanation.

4. NUMERICAL EVALUATIONS

We now present experimental evaluations with the method proposed in this paper. Data for these experiments has been collected in the acoustic chamber within the Multimedia Systems Laboratory at SMU. The wall treatments for these experiments were chosen to obtain a reverberation time of 300 ms. The microphone array chosen for the experiments employs between two and four omnidirectional lapel microphones in an approximate linear array with a nominal 4cm spacing. The chosen setup employs three loudspeakers, one of which acts as the speech source and the other two act as noise sources. For the case in which the speech source is corrupted by a single noise source, the directions of arrival (DOA) of the speech source and the noise source are approximately -30^o and 30^o, respectively, from the array normal. For the case in which two noise sources are used, the DOAs of the noise sources are -30^o, 0^o, whereas the DOA of the speech source is 30^o. All sound sources are equidistant from the array and are located 1.25m away from the array. All measurements were made using 10s of data per channel at 48kHz sampling rate and were downsampled to an 8kHz sampling rate for processing.

For each data set, the first 3s contains only noise, whereas the last 7 seconds contains speech and noise. Thus, the initial 3s and last 7s of data can be used to estimate $\{\mathbf{R}\mathrm{v}_p\}$) and $\{\mathbf{R}\mathrm{y}_p\}$) for the proposed algorithm. The algorithm was allowed to run for 100 iterations in every case with $\mu = 0.5$, after which the signal $y_1(l)$ was found to largely contain the speech source of interest. The total processing time with our MATLAB implementation in this setup varied from approximately 1.5 seconds for $\{n = 2, L = 256\}$ to 25 seconds for $\{n = 4, L = 1024\}$ on a 3.6GHz single-core Pentium PC. Since recorded speech was employed, least-squares methods were used to estimate the contributions of this speech signal before and after processing to determine initial SIRs and the SIR improvement obtained by the algorithm.

Extensive experiments were run to understand the algorithm's behavior under the following variations: 1) differing noise distributions [pink, babble, and pink+babble], 2) differing initial SIRs [from -10dB to 10dB], 3) different numbers of microphones [from two to four], and 4) different values of the filter length parameter L [from 256 to 1024]. Tables 1, 2, and 3 present the results of these evaluations, in which the SIR improvement (Final SIR - Initial SIR) is tabulated. Based on these results, we have determined the following:

1. The algorithm's performance is impressive across all SIRs

Table 1. SIR gain in dB with a single pink noise interferer; RT = 300ms

	2 Microphone System				3 Microphone System				4 Microphone System			
Initial SIR	L=256	L=512	L=768	L=1024	L=256	L=512	L=768	L=1024	L=256	L=512	L=768	L=1024
-10 dB	9.99	12.80	14.71	16.13	11.32	14.42	16.33	17.53	13.00	16.10	17.77	18.62
- 5 dB	12.00	15.10	17.21	18.80	13.32	16.72	18.80	20.05	15.19	18.52	20.18	21.14
0 dB	12.04	15.23	17.34	18.92	13.35	16.82	18.85	20.02	15.20	18.48	20.02	21.00
5 dB	11.98	15.15	17.23	18.79	13.28	16.70	18.67	19.82	15.12	18.40	19.92	20.91
10 dB	11.67	14.75	16.68	18.02	12.84	16.12	17.91	18.88	14.47	17.52	18.87	19.68

Table 2. SIR gain in dB with a single babble noise interferer; RT = 300ms

	2 Microphone System				3 Microphone System				4 Microphone System			
Initial SIR	L=256	L=512	L=768	L=1024	L=256	L=512	L=768	L=1024	L=256	L=512	L=768	L=1024
-10 dB	11.65	14.25	15.96	17.28	16.05	18.37	19.81	20.48	17.53	20.13	21.24	20.62
- 5 dB	14.46	17.49	19.40	20.92	18.38	21.17	22.90	23.85	19.97	22.98	24.37	24.15
0 dB	14.83	17.95	19.88	21.43	18.62	21.45	23.16	24.12	20.28	23.35	24.77	24.80
5 dB	15.08	18.17	20.06	21.54	18.69	21.43	23.00	23.85	20.24	23.10	24.35	24.38
10 dB	14.87	17.56	19.02	20.03	17.82	19.85	20.81	21.28	18.99	20.90	21.58	21.56

Table 3. SIR gain in dB with a babble noise interferer and a pink noise interferer; RT = 300ms

	2 Microphone System				3 Microphone System				4 Microphone System			
Initial SIR	L=256	L=512	L=768	L=1024	L=256	L=512	L=768	L=1024	L=256	L=512	L=768	L=1024
-10 dB	5.29	5.88	6.26	6,51	11.48	13.15	14.28	15.04	12.71	14.50	15.42	15.46
- 5 dB	7.79	8.46	8.85	9.12	14.39	16.39	17.71	18.53	15.61	17.75	18.80	18.94
0 dB	8.43	9.16	9.54	9.80	14.98	17.13	18.56	19.43	16.28	18.55	19.68	20.00
5 dB	8.92	9.70	10.09	10.35	15.23	17.27	18.58	19.33	16.42	18.50	19.44	19.75
10 dB	9.15	9.96	10.38	10.64	15.20	17.15	18.32	18.97	16.34	18.29	19.09	19.33

considered. For example, it provides over 20dB of SIR gain for an initial SIR of -10dB with babble noise interference for a 4-microphone array (see Table 2). This performance saturates when the initial SIR is high and when the filter length parameter L is large.

2. The algorithm performs the best under babble noise conditions. Since babble noise is non-stationary, our estimate of $\{\mathbf{R}_{\mathrm{V}p}\})$ is clearly not an exact match to the actual noise correlation statistics during the speech-plus-noise signal period. Thus, our algorithm is not highly sensitive to estimation errors in the noise correlation statistics.
3. There is a natural tradeoff between the number of microphones n and the filter length parameter L needed to achieve a given level of performance. For the same level of enhancement, a system with more microphones requires a smaller value of L.

In all cases, the enhanced speech was found to be free of any musical tones because of the inherent input-output linearity of our processing method.

5. CONCLUSIONS

This paper describes a novel method for multi-microphone speech enhancement that uses knowledge of the spatio-temporal characteristics of the noise field in an iterative procedure. The algorithm does not involve matrix inverses or large-scale matrix-vector multiplications, converges quickly, and requires little fine-tuning. Extensive numerical experiments under different scenarios show significant SIR gains for a broad range of initial SIR conditions, without introducing musical tone artifacts in the enhanced speech. The algorithm never diverged during any of our experiments. A theoretical analysis of the algorithms' convergence behavior is underway.

6. REFERENCES

[1] Y. Ephraim and H. L. Vantrees, "A signal subspace approach for speech enhancement," *IEEE Transactions Speech Audio Processing*, vol. 3, no. 4, pp. 251 – 266, Jul. 1995.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions Acoustics Speech Signal Processing*, vol. 27, no. 2, pp. 113 – 120, 1979.

[3] U. Mittal and N. Phamdo, "Signal/Noise KLT based approach for enhancing speech degraded by correlated noise," *IEEE Transactions Speech Audio Processing*, vol. 8, no. 2, pp. 159 – 167, Mar. 2000.

[4] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Transactions Speech Audio Processing*, vol. 9, no. 2, pp. 87 – 95, Feb. 2001.

[5] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Transactions Speech Audio Processing*, vol. 3, no. 6, pp. 439 – 448, Nov. 1995.

[6] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by correlated noise," *IEEE Transactions Speech Audio Processing*, vol. 11, no. 4, pp. 334 – 341, Jul. 2003.

[7] M. Brandstein and D. Ward, eds., *Microphone Arrays: Signal Processing Techniques and Applications* (New York: Springer, 2001).

[8] B. D. V. Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP. Mag.*, pp. 4 – 24, Apr. 1988.

[9] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions Signal Processing*, vol. 50, no. 9, pp. 2230 – 2244, Sep 2002.

[10] F. T. Luk, "A parallel method for computing the generalized singular value decomposition," *J. Parallel Distributed Computing*, vol. 2, no. 3, pp. 250 – 260, Aug. 1985.

[11] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. (Baltimore: John Hopkins University Press, 1996).

[12] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks : Theory and Applications* (New York: Wiley-Interscience, 1996).

[13] S. C. Douglas, "Simple adaptive algorithms for cholesky, LDL^T, QR, and eigenvalue decompositions of autocorrelation matrices for sensor array data," in *Thirty-Fifth Asilomar Conf. Signals Systems Computers*, Pacific Grove, CA, vol. 2, pp. 1134 – 1138, Nov. 2001.

[14] S. C. Douglas and M. Gupta, "Scaled natural gradient algorithms for instantaneous and convolutive blind source separation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Honolulu, HI, vol. 2, pp. 637–640, Apr. 2007.